

WS4A: a Biomedical Question and Answering System based on public Web Services and Ontologies

Miguel J. Rodrigues *, Miguel Falé, Andre Lamurias, and Francisco M. Couto

LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract. This paper describes our system, dubbed WS4A (Web Services for All), that participated in the fourth edition of the BioASQ challenge (2016). We used WS4A to perform the Question and Answering (QA) task 4b, which consisted on the retrieval of relevant concepts, documents, snippets, RDF triples, exact answers and "ideal answers" for each given question. The novelty in our approach consists on the maximum exploitation of existing web services in each step of WS4A, such as the annotation of text, and the retrieval of metadata for each annotation. The information retrieved included concept identifiers, ontologies, ancestors, and most importantly, PubMed identifiers. The paper describes the WS4A pipeline and also presents the precision, recall and f-measure values obtained in task 4b.

Keywords: natural language processing, web services, name entity recognition, information retrieval, support vector machines, semantic similarity, question and answering

1 Introduction

This paper describes our participation in the BioASQ challenge edition of 2016¹, a challenge on large-scale biomedical semantic indexing and question answering. Our participation focused on the Question and Answering (QA) task 4b, which consisted on the retrieval of relevant concepts, documents, snippets, RDF triples, exact answers and "ideal answers" for each given question. The competition was composed of two phases: in the first one the goal was to take as input a set of queries, and respond with set of the most relevant concepts, articles, snippets and RDF triples. The ontologies and terminologies from where the list of concepts were to be retrieved are: the Medical Subject Headings (MeSH); the Gene Ontology (GO); the Universal Protein Resource (UniProt); the Joint Chemical Dictionary (Jochem); the Disease Ontology (DO). In the second phase, gold responses to the questions answered of the first phase were provided, so this time, the goal was to provide a response by exploring the correct documents and snippets gathered by these experts. This challenge required the use of selected ontologies and answers in a required format according to the type of query.

* Corresponding author: mrodrigues@lasige.di.fc.ul.pt

¹ <http://www.bioasq.org/participate/challenges>

We developed the system WS4A (Web Services for (4) All), which novelty was to explore every possible option to use public web services and incorporate available domain knowledge. WS4A addressed the above tasks by first recognizing relevant terms in the query and also in the abstracts associated with it based on available Web Services. Next, the system mapped those terms to the respective concepts in ontologies and terminologies presented below. Then, WS4A compared those concepts to identify the responses that shared most concepts with the ones associated to the query. WS4A employed semantic similarity to measure how close in meaning they are even if they do not share the same exact concepts. Additionally, WS4A used Machine Learning [1] techniques to classify if an abstract is either relevant or not for the given query.

Section 2 describes the exploited Web Services by WS4A, and Section 3 explains its composing modules. In Sections 4 and 5, we present and discuss our results, along with future work.

2 Web Services

The first Web Service used was the one provided by BioPortal [2], that given a text returns the ontology concepts mentioned in it. There are some other parameters that WS4A explored, but with the purpose of better filtering the results, such as: i) longest annotation only (set to false); ii) number exclusion (set to false); iii) whole word only (set to true) and iv) synonym exclusion (set to false). The results were provided as a JSON dictionary, divided by annotation and ontology.

With the UniProt Web Services [3], we may gather PubMed IDs from protein descriptions, but to get these, we need to use another Web Service in order to obtain this description, the Whatizit² Web Service.

Whatizit [4] is an alternative to BioPortal, when it comes to identifying specific terms in sentences. The Web Service present in Whatizit identifies the words in a sentence according to a given vocabulary and as a result provides a XML with an identifier of the Uniprot database in one of its tags.

Back to UniProt, the request is made in the following format: i) the Web Service URL; ii) followed by a protein identifier; iii) ending with the available format that we chose (from our understanding, XML and HTML formats were available). For example: `http://www.uniprot.org/uniprot/P12345.xml`

When it comes to retrieving PubMed identifiers (PubMed IDs), the PubChem Web Service [5] was the last one to be used. This Web Service provided an easy interface, since it required trivial parameters and URLs. Each request was structured in the following way:

- Base URL: `http://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/`
- A word in the question
- Data Format: `xrefs/PubMedID/JSON`. Depending on the word, the result set (in XML) would be empty or not. If not, it meant that some PubMed IDs were present.

² <http://www.ebi.ac.uk/webservices/whatizit/info.jsf>

An example of the request for "oxygen" in the JSON data format would be: `https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/oxygen/xrefs/PubMedID/JSON`,

NCBI has a service, eutils [6], that provides a broad range of options. Two of them were used in order to retrieve all PubMed articles by their identifiers. The dates of these articles had to be before November 19th, 2015. The URLs are easy to identify and build, since they have a specific format, for example `http://www.ncbi.nlm.nih.gov/pubmed/23687640` where 223687640 is the PubMed ID. From eutils we used two services, one that searches for PubMed IDs with MeSH annotations for example: `http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&mindate=2014&maxdate=2015/11/19&term=+AND+%28GENES%5Bmesh%5D+%29`; and one that fetches the abstracts from the Ids retrieved from the previous URL, for example: `http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&retmode=xml&id=26580448,26580161,26575237,26577665`.

Finally, in the Answer Builder module, we use yet another Web Service where we use Linked Life Data and their SPARQL query endpoint, to generate rdf triples according to the ranked abstracts and annotations.³⁾

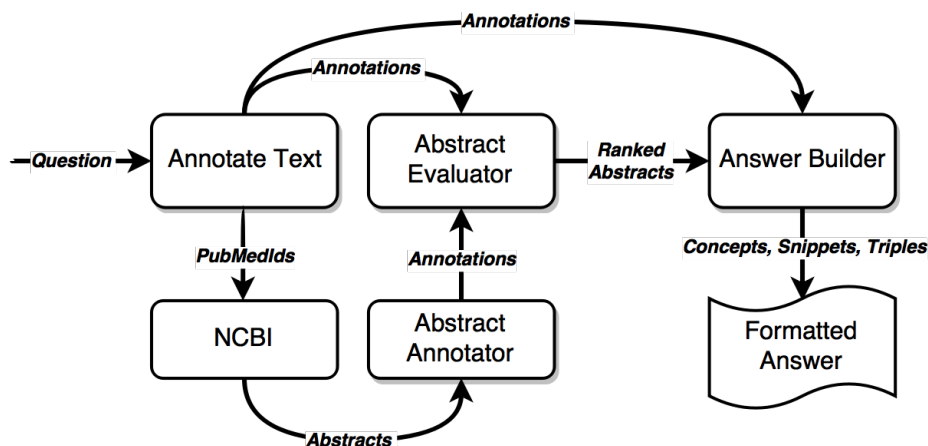


Fig. 1. The pipeline of WS4A with its main modules.

3 Pipeline

Figure 1 shows the modular nature of WS4A, where each module was designed to be as much independent from the others as possible.

³ <http://linkedlifedata.com>

We have a module, Annotate Text, where question annotations by ontologies takes place. In this module, the Web Services used are the ones described in Section 2, in particular the BioPortal web services, which returns the annotations by their respective ontologies. Following the annotation retrieval, each annotation is used to gather the respective PubMed IDs related to these annotations using the NCBI module, which also includes a local data structure containing the CheBi ontology.

After the PubMed IDs are gathered into one structure, only the ten most recent articles are used, independently of the ontology it comes from. This measure was taken just to improve temporal performance. The Abstract Annotator module comes in use: we take these 10 abstracts in order to obtain annotations from the abstracts per ontology. This module is in fact the Annotate Text module, which reuses the same functions and methods present previously, but changing the input to abstracts instead of questions.

In the Abstract Evaluator module, only the abstracts that are deemed to contain useful information are used in the response. For each abstract, the following scores are registered for each abstract:

- Jaccard score between the query and abstract annotations ⁴
- hierarchical distance score between the query and abstract annotations
- frequency between the top semantic annotations of the abstract and query
- sentence semantic similarity score ⁵

After these steps, a grade is given to the abstract using these scores. Depending on the operation being computed (training or classification of abstracts), there are two possible courses of action: approval (or disapproval) of the abstract, so it can take part in the response in case of being in the training phase; otherwise, it adds the scores to a main structure, along with the abstract to be used to generate the n-grams.

After ranking and selecting the abstracts, we use another module, Answer Builder, that selects the 10 (according to the BioASQ's rules) best valued snippets from the semantic analysis. Also, we generate the concepts required for the response. We take the initial annotated terms from the query with no regard of ontology. WS4A does the same to the annotations from the abstracts, but sums up the semantic similarity from each annotation, selecting the top 10 annotations by their score.

The following step consists of taking these top 10 annotations from the abstracts, intercept them with the original annotations. With this result, we use only the MeSH annotations, to generate the RDF triples. To filter the resulting great amount of rdf triples, we use tf-idf in order to obtain those top 10 that add more content.

The challenge provided training and test datasets that we used for our machine learning approach. In CLEF 2015 [7] among the systems that entered the BioASQ task 3b [8], we noted that OAQA [9] used supervised learning techniques to predict answer types (answer type coercion) and score candidate answers. We used Support vector machines [10,1] to classify the relevance of the abstracts to a given query. From the Abstract Evaluation module we select to work as features the four scores of the Abstract Evaluator module and the top 5000 n-grams.

⁴ <http://ag.arizona.edu/classes/rnr555/lecnotes/10.html>

⁵ <http://sujitpal.blogspot.pt/2014/12/semantic-similarity-for-short-sentences.html>

4 Results

Batch		Mean Precision		Mean Recall		Mean FMeasure		Time (sec)	
		ML	w/o ML	ML	w/o ML	ML	w/o ML	ML	w/o ML
1	Docs	0.010	0.010	0.012	0.016	0.010	0.012	16 125	8690
	Snippets	0.004	0.004	0.005	0.006	0.004	0.005		
2	Docs	0.005	0.003	0.002	0.012	0.003	0.005	11 130	8952
	Snippets	0.0	0.001	0.000	0.004	0.000	0.001		
3	Docs	0.009	0.005	0.006	0.004	0.007	0.004	10 005	8900
	Snippets	0.002	0.000	0.002	0.000	0.002	0.000		
4	Docs	0.002	0.001	0.002	0.002	0.002	0.002	11 940	8967
	Snippets	0.001	0.001	0.001	0.001	0.002	0.001		
5	Docs	0.007	0.005	0.013	0.013	0.002	0.005	9790	8358
	Snippets	0.001	0.003	0.001	0.004	0.001	0.003		

Table 1. WS4A results for all batches using the final version

Table 1 shows WS4A results for every batch and using the final version of the system. The times were obtained in a desktop computer equipped with an Intel(R) Core(TM)2 Duo CPU and 6 GB RAM.

Our results, among all participants placed on the lower half of the result table, with the exception of the fourth batch results in which we did not submit any result set. The best result F-measure scored 0.24 in average, and so the overall results were also not very high. This results reflect phase A. In phase B, we were able to achieve a top 3 result for the first batch, obtaining lower level table results for the other batches.

Comparing to last year’s results [8], WS4A was far from achieving a comparable performance mainly because we registered for this competition some weeks before the date of the first batch release and no time for tuning the system was available. Also, in this year’s competition, the type of answer wanted would already be given, so last year’s results are not strictly comparable comparing to ours.

The average time to answer a question was of around 90 seconds, with slow internet connection sometimes piking the time for obtaining an answer. W4SA was not developed to obtain highly accurate results but instead to use as much web services as possible. We demonstrated that such approach is feasible, however, we faced some issues related to availability faults in these services that hindered the progress of our system and the generation of the answers file.

We did not notice any other limitations, such as query limits (at least that we found or read about), but the servers would sometimes be unavailable for some reason for which we could never figure out the cause, and since this problem was quite random, we think it did not have anything to do with our research.

5 Conclusions

In this paper we showed the feasibility of developing a question and answering system mostly based on web services. Many of the techniques employed in WS4A have been used previously in other systems. For example, IIT [11] retrieved documents using PubMed's web services like WS4A did, and extracted snippets based on similarity between sentences of the retrieved documents and the query. LIMSI-CNRS [12] used WordNet relations (namely, synonyms) when comparing between words from the query and the answer choice, and the given short text. WS4A also used WordNet, when comparing between the words in the query and the words in the abstracts. The main difference between WS4A and the aforementioned systems, is that our system is mainly based on web services and fully explores the semantics given by the ontologies. Thus WS4A is a light system that can be easily deployed, and which is continuously updated given the extensive use of web services.

In the future, we intend to explore semantic similarity measures [13], and build a cache database in order to save and revisit results. Snippet generation can also be improved through semantic similarity, as CoMiC [14] used semantic similarity for the short text's segmentation in the Entrance Exams task 2015 (using the C99 algorithm). We also plan on improving the annotation gathering functionality, and include other sources such as DBpedia⁶ and YAGO⁷ for obtaining such annotations. YodaQA [15] resorts to DBpedia Spotlight, a service that automatically annotates DBpedia concepts from plain text.

Acknowledgments

This work was supported by FCT through funding of LaSIGE Research Unit, ref. UID/CEC/00408/2013.

References

1. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research* **12** (2011) 2825–2830
2. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., et al.: Biportal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research* (2009) gkp440
3. Consortium, U., et al.: Uniprot: a hub for protein information. *Nucleic acids research* (2014) gku989
4. Rebholz-Schuhmann, D., Arregui, M., Gaudan, S., Kirsch, H., Jimeno, A.: Text processing through web services: calling whatizit. *Bioinformatics* **24**(2) (2008) 296–298
5. Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Bryant, S.H.: Pubchem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research* (2009) gkp456

⁶ <http://dbpedia.org>

⁷ <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>

6. Sayers, E., Wheeler, D.: Building customized data pipelines using the entrez programming utilities (eutils). (2004)
7. Linda Cappellato, Nicola Ferro, G.F.J., Juan, E.S.: Preface. In: CLEF 2015. (2015)
8. Balikas, G., Kosmopoulos, A., Krithara, A., Paliouras, G., Kakadiaris, I.: Results of the bioasq tasks of the question answering lab at clef 2015. In: CLEF 2015. (2015)
9. Zi Yang, Niloy Gupta, X.S.D.X.C.Z., Nyberg, E.: Learning to answer biomedical factoid & list questions: Oaqa at bioasq 3b. In: CLEF 2015. (2015)
10. Joachims, T.: Making large scale svm learning practical. Technical report, Universität Dortmund (1999)
11. Harish Yenala, Avinash Kamineni, M.S., Chinnakotla, M.: Iiith at bioasq challenge 2015 task 3b: Bio-medical question answering system. In: CLEF 2015. (2015)
12. Gleize, M., Grau, B.: Limsi-cnrs@clef 2015: Tree edit beam search for multiple choice question answering. In: CLEF 2015. (2015)
13. Couto, F.M., Pinto, H.S.: The next generation of similarity measures that fully explore the semantics in biomedical ontologies. *Journal of bioinformatics and computational biology* **11**(05) (2013) 1371001
14. Ramon Ziai, B.R.: Comic: Exploring text segmentation and similarity in the english entrance exams task. In: CLEF 2015. (2015)
15. Petr Baudis, J.S.: Biomedical question answering using the yodaqa system: Prototype notes. In: CLEF 2015. (2015)